Bayesian Analysis (2024) **TBA**, Number TBA, pp. 1–30

Bayesian Analysis of Constrained Gaussian Processes

Hassan Maatouk[*,‡], Didier Rullière[†], and Xavier Bay[†]

Abstract. Due to their flexibility Gaussian processes are a well-known Bayesian framework for nonparametric function estimation. Integrating inequality constraints, such as monotonicity, convexity, and boundedness, into Gaussian process models significantly improves prediction accuracy and yields more realistic credible intervals in various real-world data applications. The finite-dimensional Gaussian process approximation, originally proposed in Maatouk and Bay (2017) is considered. This method involves approximating a parent GP by utilizing a finite-dimensional GP obtained through appropriate basis expansions. It satisfies interpolation conditions and handles a wide range of inequality constraints everywhere. Our contribution in this paper is threefold. First, we extend this approach to handle noisy observations and multiple, more general convex and non-convex constraints. Second, we propose new basis functions in order to extend the smoothness of sample paths to differentiability of class C^p, for any $p \geq 1$. Third, we examine its behavior in specific scenarios such as monotonicity with flat regions and boundedness near lower and/or upper bounds. In that case, we show that, unlike the Maximum a posteriori (MAP) estimate, the mean a posteriori (mAP) estimate fails to capture flat regions. To address this issue, we propose incorporating multiple constraints, such as monotonicity with bounded slope constraints. According to the theoretical convergence and based on a variety of numerical experiments, the MAP estimate behaves well and outperforms the mAP estimate in terms of prediction accuracy. The performance of the proposed approach is confirmed through synthetic and real-world data studies.

MSC2020 subject classifications: Primary 62G05, 62G08; secondary 62G09.

Keywords: Gaussian processes, multiple shape constraints, convex and non-convex constraints, flat region, MAP estimate, HMC sampler.

1 Introduction

Gaussian processes (GPs) are a well-known nonparametric Bayesian framework for function estimation. They are widely used in many fields, such as computer science, physics, biology, engineering, and finance (Rasmussen and Williams, 2006). GP models are based on defining a prior distribution over function spaces. In general, a GP is characterized by its mean and covariance functions. The flexibility of GPs is attributed to their covariance function, which enables incorporating prior information, such as smoothness, stationar-

[*]CY Tech, CY Cergy Paris Université, Laboratoire AGM, Site du Parc, 95011 Cergy-Pontoise, France, hmk@cy-tech.fr

[†]Mines Saint-Étienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, F-42023 Saint-Étienne France, drulliere@emse.fr; bay@emse.fr

[‡]Corresponding author.

 https://doi.org/10.1214/24-BA1429

ity, sparsity, and derivative constraints (Cramer and Leadbetter, 1967; Rasmussen and Williams, 2006).

However, it is crucial to acknowledge the nuanced limitations that GP encounter, particularly when confronted with shape constraints such as monotonicity, boundedness, and convexity, see for example, Golchi et al. (2015); Lenk and Choi (2017); Ray et al. (2020); Wang and Berger (2016). The challenge lies in their relative inflexibility when handling such constraints, leading to a potential drawback in scenarios that demand precise adherence to inequality conditions. Several real-world cases where the data suggest that the underlying function satisfies specific inequality constraints are presented in physics (Zhou et al., 2019) and econometrics (Chataigner et al., 2021; Cousin et al., 2022, 2016; Crépey and Dixon, 2020). Addressing such constraints remains an active area of research to enhance the applicability of GP models in diverse settings.

Including inequality constraints into a GP model improves its prediction accuracy and provides more realistic confidence intervals (Chipman et al., 2022; López-Lopera et al., 2018; Lin and Dunson, 2014; Riihimäki and Vehtari, 2010; Ustyuzhaninov et al., 2020; Zhou et al., 2022). The authors in Swiler et al. (2020) provide an overview and survey of various strategies for incorporating shape constraints into a GP. In Tran et al. (2023), constrained GPs have been employed to estimate probability density functions (pdf). Recently, an efficient transformed GP (ETGP) method using normalizing flow has been developed in Maroñas et al. (2021); Maroñas and Hernández-Lobato (2023). The ETGP approach involves transforming a single sample from a GP using an invertible transformation, which made input-dependent. The authors in Maroñas and Hernández-Lobato (2023) demonstrate its efficiency in multi-class classification tasks based on GPs, offering a reduced computational cost. In Maroñas et al. (2021), the authors outline the computational performance and the ability of transformed GP to incorporate prior knowledge such as boundedness constraints. In the present paper, the finite-dimensional GP approximation proposed in Maatouk and Bay (2017) is considered, where various shape constraints such as monotonicity, convexity, and boundedness are satisfied everywhere. To the best of our knowledge, it is the only model in the literature capable of dealing with a variety of shape constraints, whether applied alone, together, or sequentially. The main idea is to approximate the samples of the parent GP by representing them in a finite-dimensional space of functions using appropriate basis expansions. These basis functions possess attractive properties not necessary shared by other basis such as Bernstein polynomials (Curtis and Ghosh, 2011), regression splines (Cai and Dunson, 2007; Meyer et al., 2011), and restricted splines (Shively et al., 2011). Various restrictions like monotonicity, convexity, and boundedness are *equivalently* translated into linear inequality constraints on the basis coefficients. The performance of this approach has been demonstrated through several real-world data applications (Cousin et al., 2022, 2016; Williams et al., 2023; Zhou et al., 2019). The asymptotic properties have been investigated in Bay et al. (2016); Grammont et al. (2022). The generalization of the well-known Kimeldorf-Wahba correspondence (Kimeldorf and Wahba, 1970) between Bayesian estimation on stochastic processes and splines for the constrained cases has been established.

The present paper extends the work of Maatouk and Bay (2017) in three directions:

- We handle both noisy observations and multiple and more general convex and non-convex constraints (such as boundedness within a non-convex set).
- We propose new basis functions to enhance the smoothness of the sample paths, achieving C^p class differentiability[1], for any $p \geq 1$.
- We investigate the behavior of this approach in challenging situations, such as monotonicity with flat regions or boundedness where the underlying function is close to the lower or upper bounds. To address this issue, we propose adding multiple constraints, such as monotonicity with bounded slope constraints. This leads to correction of the posterior distribution's behavior and the convergence of the mAP estimate towards the MAP estimate.

This article is structured as follows. In Section 2, GP regression is briefly reviewed. In Section 3, following the finite-dimensional GP approximation from Maatouk and Bay (2017), we propose a general formulation for linear inequality constraints that is capable of handling both convex and non-convex constraints. Section 4 presents the new basis functions in order to generalize the smoothness of sample paths to differentiability of class C^p, $p \geq 1$. Additionally, the asymptotic properties of the MAP estimate are investigated. Section 5 demonstrates the efficiency of the proposed framework through applications using real-world data.

2 Gaussian process regression review

A GP, namely $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$, is characterized by its mean function μ and covariance function k, i.e., $Z \sim \mathcal{GP}(\mu, k)$ (Rasmussen and Williams, 2006). It can be written as

$$Z(\boldsymbol{x}) = \mu(\boldsymbol{x}) + Y(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{R}^d,$$

where $(Y(\boldsymbol{x}))$ is a zero-mean GP with covariance function k, i.e., $Y \sim \mathcal{GP}(0, k)$, with

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) = \mathrm{E}[Y(\boldsymbol{x})Y(\boldsymbol{x}')], \quad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d.$$

Given a dataset of size n, namely, $D = \{(\boldsymbol{x}_i, y_i), i = 1, \ldots, n\}$, where $\boldsymbol{x}_i$ denotes an input vector of dimension d and y_i denotes a scalar output. The input vectors $\{\boldsymbol{x}_i\}$ form the $n \times d$ design matrix $\mathbb{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]^\top$ and the outputs $\{y_i\}$ form the output vector $\boldsymbol{y} = [y_1, \ldots, y_n]^\top$ called data. Thus, the dataset can be written as $D = \{(\mathbb{X}, \boldsymbol{y})\}$. The following regression problem is considered

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_{\text{noise}}), \tag{2.1}$$

where $i = 1, \ldots, n$, f is an unknown latent function that generates the data, and ϵ_i is an additive independent identically distributed (i.i.d.) zero-mean Gaussian noise with constant variance σ^2_{noise}. A GP prior distribution on the underlying function f

[1] C^p represents the set of functions that have continuous derivatives up to the p-th order.

is assumed. Conditionally on $\boldsymbol{y} = [y_1, \ldots, y_n]^\top$, the conditional process remains a GP (Rasmussen and Williams, 2006)

$$\{Y(\cdot)|\boldsymbol{y}\} \sim \mathcal{GP}\left(\mu_c(\cdot), c(\cdot,\cdot)\right), \tag{2.2}$$

where the conditional mean function μ_c and covariance function c are given as follows:

$$\begin{aligned} \mu_c(\boldsymbol{x}) &= \mathrm{E}\left[Y(\boldsymbol{x})|\boldsymbol{y}\right] = k(\boldsymbol{x},\mathbb{X})^\top \left(k(\mathbb{X},\mathbb{X}) + \sigma^2_{\text{noise}}\mathbf{I}_n\right)^{-1}\boldsymbol{y}; \\ c(\boldsymbol{x},\boldsymbol{x}') &= k(\boldsymbol{x},\boldsymbol{x}') - k(\boldsymbol{x},\mathbb{X})^\top (k(\mathbb{X},\mathbb{X}) + \sigma^2_{\text{noise}}\mathbf{I}_n)^{-1} k(\boldsymbol{x}',\mathbb{X}); \end{aligned} \tag{2.3}$$

with $\mathbf{I}_n$ the $n \times n$ identity matrix. Let us recall that $k(\mathbb{X},\mathbb{X})$ is the covariance matrix of $Y(\mathbb{X})$ of dimension $n \times n$, and $k(\boldsymbol{x},\mathbb{X}) = [k(\boldsymbol{x},\boldsymbol{x}_1), \ldots, k(\boldsymbol{x},\boldsymbol{x}_n)]^\top$ is the vector of covariance between $Y(\boldsymbol{x})$ and $Y(\mathbb{X})$ of dimension n.

3 Constrained Gaussian processes

3.1 C^0 approximation with Model (M_h)

In this section, the finite-dimensional GP approximation proposed in Maatouk and Bay (2017) is considered. Without loss of generality, let Y be a zero-mean GP with covariance function k, i.e., $Y \sim \mathcal{GP}(0,k)$. We first focus on the one-dimensional input case. The methodology is later extended to handle multidimensional input spaces (refer to Section 3.4 below). Let $\mathcal{D}$ be a compact set in $\mathbb{R}$. For simplicity, we suppose that $\mathcal{D}$ is the unit interval $[0,1]$. Let $t_j = (j-1)\times\Delta_N$, $j \in \{1,\ldots,N\}$ be a sequence of N equally spaced knots on $\mathcal{D}$, with a spacing of $\Delta_N = 1/(N-1)$. Let us mention that the methodology developed in this paper is applicable for non-uniform discretization of $\mathcal{D}$ (see the left panel of Figure 1). Let us define the three basis functions proposed in Maatouk and Bay (2017), which will be used in three different models, M_h, M_ϕ, and M_φ, in the present paper. These functions are given by

$$h_j(x) := h\left(\frac{x-t_j}{\Delta_N}\right), \quad \phi_j(x) := \int_0^x h_j(t)\mathrm{d}t, \quad \varphi_j(x) := \int_0^x\int_0^t h_j(u)\mathrm{d}u\mathrm{d}t, \tag{3.1}$$

for $j \in \{1,\ldots,N\}$, where $h(x) := (1-|x|)\mathbf{1}_{[-1,1]}(x)$ is the *hat* function on $[-1,1]$. The *hat* functions $\{h_j\}$ admit two nice properties. First, the value of any *hat* function at any knot is equal to Kronecker's delta function (i.e., $h_j(t_l) = \delta_{j,l}$), where $\delta_{j,l}$ is equal to one when $j = l$ and zero otherwise. Second, for any $x \in \mathcal{D}$, we have $\sum_{j=1}^N h_j(x) = 1$. The second property is used in the proof of Lemma 1. As mentioned in Maatouk and Bay (2017), any continuous function $f : \mathcal{D} \to \mathbb{R}$, that is, $f \in C^0(\mathcal{D},\mathbb{R})$ can be approximated by a piecewise linear interpolating between the function values at the knots $\{t_j\}$,

$$\tilde{f}_N(x) = \sum_{j=1}^N f(t_j)h_j(x), \quad \forall x \in \mathcal{D}.$$

Let us recall the following well known result.

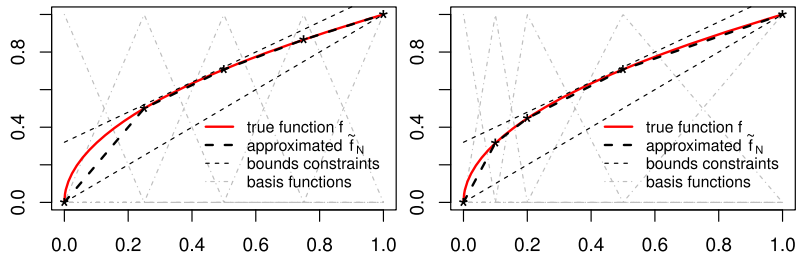

Figure 1: An example of approximating a monotone (nondecreasing) and bounded function f (red solid curve) using a piecewise linear interpolating function $\tilde{f}_N$ (black dashed curve). A uniform (resp. non-uniform) subdivision is used with $N = 5$ *hat* functions and knots in the left (resp. right) panel. The gray triangles represent the *hat* functions, while the black dashed thin lines denote the lower and upper bounds constraints.

Lemma 1 (Uniform convergence C^0)**.** *Let f be a continuous function on $\mathcal{D}$, then, the piecewise linear interpolating function $\tilde{f}_N(\cdot) = \sum_{j=1}^{N} f(t_j)h_j(\cdot)$ converges uniformly to f when N tends to infinity.*

Figure 1 shows the deterministic function $f(x) = \sqrt{x}$ (red solid curve) which verifies the monotonicity (nondecreasing) and boundedness constraints. This function is approximated by the piecewise linear interpolating function $\tilde{f}_N$ (black dashed curve) using either a uniform subdivision with $N = 5$ *hat* basis functions (left panel), or a non-uniform subdivision (right panel). The black dashed thin lines represent the lower and upper bound constraints, while the gray triangles represent the *hat* basis functions $\{h_j\}$. The black stars represent the values of the true and approximated functions at the knots $\{t_j\}$. Let us mention that the true function grows rapidly on $[0, 0.3]$. As a result, a finer discretization with $N = 3$ was used only for this interval, while $N = 2$ was used for the interval $[0.3, 1]$. This shows that a suitable subdivision can improve the accuracy of the approximation and reduce the number of knots. This also reduces the complexity of the sampling process when using an efficient HMC sampler to approximate the posterior distribution, as described in detail in Section 3.2.

If no additional smoothness assumptions are required, the first model is

$$Y^N(x) := \sum_{j=1}^{N} Y(t_j)h_j(x) = \sum_{j=1}^{N} \xi_j h_j(x), \quad x \in \mathcal{D}, \tag{M_h}$$

where we denote $\xi_j = Y(t_j)$. Since Y is a zero-mean GP with covariance function k, then, the vector $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_N]^\top$ is also zero-mean and Gaussian with covariance matrix $\boldsymbol{\Gamma}$, i.e., $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Gamma})$, where

$$\boldsymbol{\Gamma}_{j,l} = \mathrm{Cov}(Y(t_j), Y(t_l)) = k(t_j, t_l), \quad j, l \in \{1, \ldots, N\}, \tag{3.2}$$

and $\mathbf{0}_N = [0, \ldots, 0]^\top$ is the N-dimensional zero vector. Furthermore, the coefficients $\{\xi_j\}$ can be interpreted as the values of the original GP $(Y(x))$ evaluated at the knots

$\{t_j\}$. Let $\mathcal{C}$ be the convex set of functions that verify some inequality constraints, such as monotonicity, convexity, and boundedness. The non-convex case will be investigated later in this section. For instance,

$$\mathcal{C} = \begin{cases} \mathcal{C}_b := \left\{ f \in C^0(\mathcal{D}, \mathbb{R}) \text{ s.t. } \ell \leq f(x) \leq u, \ \forall x \in \mathcal{D} \right\} \\ \mathcal{C}_m := \left\{ f \in C^0(\mathcal{D}, \mathbb{R}) \text{ s.t. } f(x) \leq f(y), \ \forall x \leq y \in \mathcal{D} \right\} \\ \mathcal{C}_c := \left\{ f \in C^0(\mathcal{D}, \mathbb{R}) \text{ s.t. } \frac{f(y)-f(x)}{y-x} \leq \frac{f(z)-f(y)}{z-y}, \ \forall x < y < z \in \mathcal{D} \right\} \end{cases} \tag{3.3}$$

which corresponds to boundedness, monotonicity, and convexity constraints respectively, where the constants ℓ and u represent the lower and upper bounds, respectively, and where $C^0(\mathcal{D}, \mathbb{R})$ is the set of continuous functions from $\mathcal{D}$ to $\mathbb{R}$. Our aim is to compute the posterior distribution of Y^N such that $Y^N \in \mathcal{C}$. The authors in Maatouk and Bay (2017) have shown the advantage of using the *hat* function as in Model (M_h) and more generally the basis functions defined in (3.1). They demonstrated that satisfying an infinite number of inequality constraints on the process $Y^N \in \mathcal{C}$ is *equivalent* to satisfying a finite number of linear inequality constraints on the coefficient vector $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_N]^\top$. To be more precise, for many various choices of $\mathcal{C}$, we have

$$Y^N \in \mathcal{C} \quad \Leftrightarrow \quad \boldsymbol{\xi} \in \mathcal{E}, \tag{3.4}$$

where $\mathcal{E}$ is a convex set of $\mathbb{R}^N$. For the inequality constraints given in (3.3), we obtain

$$\mathcal{E} = \begin{cases} \mathcal{E}_b := \left\{ z \in \mathbb{R}^N : \ \ell \leq z_j \leq u, \ \forall j = 1, \ldots, N \right\} \\ \mathcal{E}_m := \left\{ z \in \mathbb{R}^N : \ z_{j-1} \leq z_j, \ \forall j = 2, \ldots, N \right\} \\ \mathcal{E}_c := \left\{ z \in \mathbb{R}^N : \ \dfrac{z_{j-1} - z_{j-2}}{t_{j-1} - t_{j-2}} \leq \dfrac{z_j - z_{j-1}}{t_j - t_{j-1}}, \ \forall j = 3, \ldots, N \right\} \end{cases} \tag{3.5}$$

which corresponds to boundedness, monotonicity, and convexity constraints respectively. The ability of these bases $\{h_j\}$ to express different constraints *equivalently* as linear restrictions on the vector of coefficients $\boldsymbol{\xi}$ is a desirable feature that may not be present in other bases such as Bernstein polynomials (Curtis and Ghosh, 2011), regression splines (Cai and Dunson, 2007; Meyer et al., 2011), and restricted splines (Shively et al., 2011). In the following sections, we also demonstrate the attractiveness of the other basis functions defined in (3.1). Furthermore, new and smoother basis functions are given below (cf., Section 4). In the present section, we focus exclusively on Model (M_h) using the *hat* functions $\{h_j\}$. Note that the linear restrictions on the coefficients vector $\boldsymbol{\xi}$ in Eq. (3.5) can be expressed in matrix form as follows:

$$\boldsymbol{\xi} \in \mathcal{E} \quad \Leftrightarrow \quad \boldsymbol{\xi} \in \mathbb{R}^N \text{ s.t. } \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}, \tag{3.6}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times N}$ is the matrix of constraints, and $\boldsymbol{\ell}$ and $\boldsymbol{u}$ are lower and upper bounds vectors, respectively. For instance, when $\boldsymbol{\xi} \in \mathcal{E}_m$, we get

$$\boldsymbol{\Lambda}_{i,j} = \begin{cases} -1 & \text{if } j = i & \text{for any } i = 1, \ldots, N-1; \\ 1 & \text{if } j = i+1 & \text{for any } i = 1, \ldots, N-1; \\ 0 & \text{otherwise;} \end{cases}$$

and $\boldsymbol{\ell} = [0, \ldots, 0]^\top \in \mathbb{R}^m$ and $\boldsymbol{u}$ is the vector with components $+\infty$. In that case, we get $m = N - 1$ linear inequality constraints on the coefficients vector $\boldsymbol{\xi}$.

The authors of Maatouk and Bay (2017) demonstrated that under the representation in Model (M_h):

- Y^N is a finite-dimensional GP with covariance function

$$k_N(x,x') = \boldsymbol{h}(x)^\top \boldsymbol{\Gamma} \boldsymbol{h}(x'), \quad \forall x, x' \in \mathcal{D},$$

  where $\boldsymbol{h}(x) = [h_1(x), \ldots, h_N(x)]^\top$.

- Y^N converges uniformly to Y when N tends to infinity (with probability one).

Let us give the following results concerning Model (M_h).

Proposition 1 (Multiple constraints (M_h))**.** *Suppose that Y^N is defined as in Model (M_h).*

- *Boundedness in a convex region: Let $\mathcal{C}_{bc}$ be a set of continuous functions on $\mathcal{D}$ bounded between two functions f_ℓ and f_u such that the region between the lower bound function f_ℓ and the upper bound function f_u is convex.*[2] *Then, $Y^N \in \mathcal{C}_{cb}$ if and only if $f_\ell(t_j) \leq \xi_j \leq f_u(t_j)$, for any $j \in \{1, \ldots, N\}$.*

- *Convexity: Y^N is convex (i.e., $Y^N \in \mathcal{C}_c$) if and only if*

$$\frac{\xi_{j+1} - \xi_j}{t_{j+1} - t_j} \leq \frac{\xi_{j+2} - \xi_{j+1}}{t_{j+2} - t_{j+1}}, \quad j \in \{1, \ldots, N-2\}. \tag{3.7}$$

  In our case, $t_{j+1} - t_j = \Delta_N = 1/(N-1)$, for any $j \in \{1, \ldots, N-1\}$. Thus, inequalities (3.7) are equivalent to

$$\xi_{j+2} - 2\xi_{j+1} + \xi_j \geq 0, \quad j \in \{1, \ldots, N-2\},$$

  which in turn is equivalent to $\boldsymbol{\Lambda}\boldsymbol{\xi} \geq \mathbf{0}_{N-2}$ according to the notation in (3.6), with $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_N]^\top$. The matrix of constraints $\boldsymbol{\Lambda}$ is given by

$$\boldsymbol{\Lambda}_{i,j} = \begin{cases} 1 & \text{if } j = i \quad \text{for any } i = 1, \ldots, N-2; \\ -2 & \text{if } j = i+1 \quad \text{for any } i = 1, \ldots, N-2; \\ 1 & \text{if } j = i+2 \quad \text{for any } i = 1, \ldots, N-2; \\ 0 & \text{otherwise.} \end{cases}$$

- *Multiple constraints: For example, Y^N is nondecreasing and bounded in a convex region (i.e., $Y^N(x) \in \mathcal{C}_m \cap \mathcal{C}_{bc}$) if and only if*

$$\begin{cases} \xi_j \leq \xi_{j+1} & j \in \{1, \ldots, N-1\}; \\ f_\ell(t_j) \leq \xi_j \leq f_u(t_j) & j \in \{1, \ldots, N\}; \end{cases} \tag{3.8}$$

  where f_ℓ and f_u are the lower and upper bounds functions. Let us mention that if f_ℓ and f_u are constants, for example $f_\ell(x) = a < b = f_u(x)$, for any $x \in \mathcal{D}$, where $a, b \in \mathbb{R}$, then the linear constraints (3.8) become $a \leq \xi_1 \leq \ldots \leq \xi_N \leq b$. This leads to only $m = N+1$ linear constraints on $\boldsymbol{\xi}$ according to (3.6).

[2] A set $\mathcal{C}$ is convex if and only if for any $x, y \in \mathcal{C}$, $(1-t)x + ty \in \mathcal{C}$, where $t \in]0,1[$.

Remark 1.

- *The linear inequality constraints on the coefficients* $f_\ell(t_j) \leq \xi_j \leq f_u(t_j)$*, for any* $j \in \{1, \ldots, N\}$ *in Proposition 1 can be written in matrix form as in (3.6), where* $\boldsymbol{\Lambda} = \boldsymbol{I}_N$*,* $\boldsymbol{\ell} = [f_\ell(t_1), \ldots, f_\ell(t_N)]^\top$*, and* $\boldsymbol{u} = [f_u(t_1), \ldots, f_u(t_N)]^\top$*.*
- *The three constraints introduced in Proposition 1 (boundedness, convexity, and monotonicity) can be imposed together.*

The result of Proposition 1 (boundedness constraints) remains valid if the region between the lower and upper bounds functions is non-convex, as long as it can be decomposed into convex sets at nonoverlapping subdomains (cf., Proposition 2 and Figures 2 and 4 below).

Proposition 2 (Boundedness in a non-convex region)**.** *Suppose that* $\mathcal{C}_{bnc}$ *is a set of continuous functions on* $\mathcal{D}$ *bounded between two functions* f_ℓ *and* f_u *such that the region between the lower bound function* f_ℓ *and the upper bound function* f_u *is non-convex. Suppose also that this non-convex region can be decomposed into convex regions on* Q *nonoverlapping input subdomains* $\mathcal{D}_r$*,* $r \in \{1, \ldots, Q\}$*. If the intersection extremities* $t_{1,N_1}, \ldots, t_{Q-1,N_{Q-1}}$ *of each subdomain are elements of the subdivision of* $\mathcal{D}$*, then* Y^N *is in* $\mathcal{C}_{bnc}$ *if and only if* $f_\ell(t_{r,j_r}) \leq \xi_{r,j_r} \leq f_u(t_{r,j_r})$*, for all* $j_r \in \{1, \ldots, N_r\}$ *and* $r = \{1, \ldots, Q\}$*. These linear inequalities on* ξ_{r,j_r} *can be expressed in matrix form as follows:*

$$\boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u},$$

where $\boldsymbol{\ell} = [\boldsymbol{\ell}_1^\top, \ldots, \boldsymbol{\ell}_Q^\top]^\top$*,* $\boldsymbol{u} = [\boldsymbol{u}_1^\top, \ldots, \boldsymbol{u}_Q^\top]^\top$*, and* $\boldsymbol{\Lambda} = [\boldsymbol{\Lambda}_1^\top, \ldots, \boldsymbol{\Lambda}_Q^\top]^\top$ *(block diagonal matrix), with for example,* $\boldsymbol{\ell}_1 = [f_\ell(t_{1,1}), \ldots, f_\ell(t_{1,N_1})]^\top$*.*

3.2 Constrained Gaussian process with noisy observations

In this section, we consider the finite-dimensional GP approximation defined in (M_h) given both noisy observations and inequality constraints:

$$Y^N(x) = \sum_{j=1}^{N} \xi_j h_j(x) \quad \text{s.t.} \quad \begin{cases} Y^N(x_i) + \epsilon_i = y_i & \text{(noisy observations)}, \\ Y^N \in \mathcal{C} & \text{(inequality constraints)}, \end{cases} \tag{3.9}$$

where $x_i \in \mathcal{D}$ is the design point, $y_i \in \mathbb{R}$ is the data and $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_{\text{noise}})$, with σ^2_{noise} the noise variance. Given a set of design points $\mathbb{X} = [x_1, \ldots, x_n]^\top \in \mathcal{D}^n$, the noisy observations can be written in matrix form as follows:

$$\mathbf{H}\boldsymbol{\xi} + \boldsymbol{\epsilon} = \boldsymbol{y},$$

where $\boldsymbol{y} = [y_1, \ldots, y_n]^\top$ is the vector of data, $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]^\top$ is the noise Gaussian vector, and $\mathbf{H}$ is the $n \times N$ design matrix defined by $\mathbf{H}_{i,j} := h_j(x_i)$. Following the strategy in Maatouk (2022) and the equivalent in (3.4), the conditional distribution of

Y^N given both noisy observations $\{Y^N(\mathbb{X})+\boldsymbol{\epsilon}=\boldsymbol{y}\}$ and inequality constraints $\{Y^N \in \mathcal{C}\}$ can be obtained from the conditional distribution of $\boldsymbol{\xi}$ given $\{\mathbf{H}\boldsymbol{\xi}+\boldsymbol{\epsilon}=\boldsymbol{y}\}$ and $\{\boldsymbol{\xi} \in \mathcal{E}\}$

$$\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Gamma}) \quad \text{s.t.} \quad \begin{cases} \mathbf{H}\boldsymbol{\xi}+\boldsymbol{\epsilon}=\boldsymbol{y} & \text{(noisy observations)} \\ \boldsymbol{\xi} \in \mathcal{E} & \text{(linear inequality constraints)} \end{cases} \tag{3.10}$$

Hereafter, the linear inequality constraints $\{\boldsymbol{\xi} \in \mathcal{E}\}$ is reformulated as $\boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}$, where $\boldsymbol{\ell}$ and $\boldsymbol{u}$ are the lower and upper bounds vectors of dimension m. Now, we will explain the procedure for sampling from the posterior distribution as stated in (3.10). Since $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Gamma})$, then $\mathbf{H}\boldsymbol{\xi}+\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_N, \mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^\top + \sigma^2_{\text{noise}}\mathbf{I}_n)$. Under only noisy observations, the conditional distribution of $\boldsymbol{\xi}$ is a multivariate normal (MVN), see Maatouk et al. (2022); Rasmussen and Williams (2006):

$$\{\boldsymbol{\xi}|\mathbf{H}\boldsymbol{\xi}+\boldsymbol{\epsilon}=\boldsymbol{y}\} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \text{where}$$

$$\begin{cases} \boldsymbol{\mu} &= (\mathbf{H}\boldsymbol{\Gamma})^\top(\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^\top + \sigma^2_{\text{noise}}\mathbf{I}_n)^{-1}\boldsymbol{y} \\ \boldsymbol{\Sigma} &= \boldsymbol{\Gamma} - (\mathbf{H}\boldsymbol{\Gamma})^\top(\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^\top + \sigma^2_{\text{noise}}\mathbf{I}_n)^{-1}\mathbf{H}\boldsymbol{\Gamma} \end{cases} \tag{3.11}$$

with $\mathbf{I}_n$ the $n \times n$ identity matrix. Note that this problem is called *hyperplane-truncated* MVN distribution (Cong et al., 2017; Maatouk et al., 2022, 2023b,a). The consideration of noisy observations in (3.9) has a *relaxing* effect on the interpolation conditions, as the number of knots and basis functions N does not need to be larger than the size n of the samples (condition required in Maatouk and Bay (2017) for the interpolation of noise-free observations called *degree of freedom*). This leads to less restrictive sample spaces and less expensive MCMC samplers when $N \ll n$ as it is performed on $\mathbb{R}^N$ and independent of the number of observations n. Additionally, it should be mentioned that, unlike interpolation with noise-free observations, the given data does not need to satisfy inequality constraints (see, for example, Figure 3). The posterior distribution (3.10) is the following truncated MVN distribution:

$$\{\boldsymbol{\xi}|\mathbf{H}\boldsymbol{\xi}=\boldsymbol{y}+\boldsymbol{\epsilon}, \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\} \sim \mathcal{N}_{\mathcal{T}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\ell}, \boldsymbol{u}), \tag{3.12}$$

where $\mathcal{N}_{\mathcal{T}}(\boldsymbol{m}, \boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b})$ is the truncated MVN distribution with mean vector $\boldsymbol{m}$, covariance matrix $\boldsymbol{C}$, and lower and upper bounds constraints $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively. Recently, several efficient MCMC algorithms have been proposed to approximate the truncated posterior distribution (3.12), such as Gibbs sampling (Taylor and Benjamini, 2016), Metropolis-Hastings (Murphy, 2018), HMC (Pakman and Paninski, 2014) and the minimax tilting method accept-reject sampler (Botev, 2017). In the present paper, the fast HMC sampler developed in Pakman and Paninski (2014) and implemented in the `R package` *tmg* is used.

Let us mention that the posterior mode $\boldsymbol{\mu}^*$ corresponds to the maximum of the posterior pdf

$$\boldsymbol{\mu}^* := \underset{\substack{\boldsymbol{z} \in \mathbb{R}^N \\ \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{z} \leq \boldsymbol{u}}}{\arg\max} \left\{ -[\boldsymbol{z}-\boldsymbol{\mu}]^\top \boldsymbol{\Sigma}^{-1}[\boldsymbol{z}-\boldsymbol{\mu}] \right\}.$$

This problem is equivalent to the following quadratic optimization problem subject to convex constraints (Boyd and Vandenberghe, 2004; Goldfarb and Idnani, 1983)

$$\boldsymbol{\mu}^* := \underset{\substack{\boldsymbol{z} \in \mathbb{R}^N \\ \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{z} \leq \boldsymbol{u}}}{\arg\min} \left\{ \boldsymbol{z}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{z} - 2\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{z} \right\}, \tag{3.13}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given in (3.11). In the numerical examples presented in this paper, we use the solve.QP function from the `R package` *quadprog* to calculate the posterior mode $\boldsymbol{\mu}^*$. It can then serve as a suitable starting point for the HMC sampler.

Before presenting numerical examples of the proposed approach for various types of inequality constraints, we define the following two estimators.

Definition 1 (MAP estimate)**.** *The Maximum a posteriori (MAP) estimate of* Y^N *conditionally on inequality constraints and noisy observations is defined as*

$$M^N(x) := \sum_{j=1}^{N} \mu_j^* h_j(x) = \boldsymbol{h}(x)^\top \boldsymbol{\mu}^*, \quad x \in \mathcal{D},$$

where $\boldsymbol{\mu}^* = [\mu_1^*, \ldots, \mu_N^*]^\top \in \mathbb{R}^N$ *is the posterior mode computed by* (3.13) *and* $\boldsymbol{h}(x) = [h_1(x), \ldots, h_N(x)]^\top \in \mathbb{R}^N$.

Let us provide some comments: the MAP estimate in Algorithm 1 is independent of the sampling process. It is determined only by solving a quadratic optimization problem with linear inequality constraints (3.13). Furthermore, it has been shown that the MAP estimate M^N converges to the optimization spline problem in both noise and noise-free observation cases when using the *hat* basis functions, i.e., Model (M_h), see Bay et al. (2016); Grammont et al. (2022). These two results can be seen as a generalization of the correspondence established in Kimeldorf and Wahba (1970) between Bayesian estimation on stochastic processes and smoothing by splines.

Definition 2 (mAP estimate)**.** *The mean a posteriori (mAP) estimate of* Y^N *conditionally on inequality constraints and noisy observations is defined as*

$$m^N(x) := \mathrm{E}\left[Y^N(x) \middle| Y^N(\mathbb{X}) + \boldsymbol{\epsilon} = \boldsymbol{y}, Y^N \in \mathcal{C}\right] = \boldsymbol{h}(x)^\top \overline{\boldsymbol{\mu}},$$

where $\overline{\boldsymbol{\mu}} := \mathrm{E}\left[\boldsymbol{\xi} | \boldsymbol{H}\boldsymbol{\xi} + \boldsymbol{\epsilon} = \boldsymbol{y}, \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\right]$ *is the posterior mean which is computed from simulations and* $\boldsymbol{h}(x) = [h_1(x), \ldots, h_N(x)]^\top \in \mathbb{R}^N$.

3.3 Illustrative examples of Model (M_h)

In the section, the performance of the MAP estimate using Model (M_h) is highlighted, and the flexibility of Model (M_h) to incorporate multiple types of convex and non-convex inequality constraints is shown. Other illustrative examples, such as those involving combined monotonicity and boundedness constraints applied both concurrently and sequentially, have been deferred to the supplementary material (Section 2) Maatouk et al. (2024).

Example 1 (Boundedness in a convex set)**.** *We consider the function* $f_1(x) = 0.8x \sin(5x)$ *for any* $x \in \mathcal{D}$. *This function is bounded on* $\mathcal{D}$ *between convex and concave functions* f_ℓ *and* f_u *respectively:*

$$f_\ell(x) = (x - 0.5)^2 - 1.2 \quad \textit{and} \quad f_u(x) = -(x - 0.5)^2 + 0.3.$$

This function is slightly flat and close to the upper bound function f_u *on* $[0, 0.5]$.

To avoid the possibility of overfitting (Neelon and Dunson, 2004), we plot in Figure 2, the mean squared prediction error (MSPE) for both the MAP and mAP estimates as a function of the number of basis functions N*, considering only the case with boundedness constraints. A dataset of size* $n = 70$ *generated from (2.1) using* f_1 *and a true* $\sigma_{noise} = 0.5$ *is randomly split into training set of size* 50 *and testing set of size* 20*. This numerical analysis suggests that a relatively "small" number of basis functions yields a "satisfactory" approximation. This effectiveness is attributed to the fact that the bounded function* f_1 *exhibits minimal variation across its entire domain* $\mathcal{D}$*. It is worth noting the MAP estimate outperforms the mAP estimate in terms of prediction accuracy, as it yields the smallest MSPE for any value of* N*.*

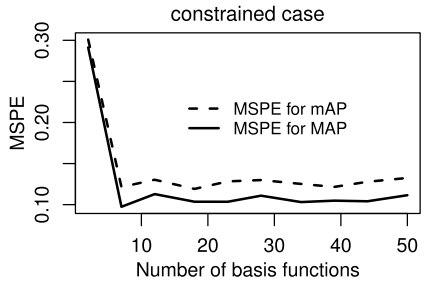

Figure 2: The average MSPE over 25 replicates as a function of the dimension N is shown using the MAP and mAP estimates from Model (M_h) under boundedness.

Figure 3 illustrates the finite-dimensional GP approximation from Model (M_h) *with (right panel) and without (left panel) boundedness constraints. We use the Matérn covariance function with a smoothness parameter* $\nu = 5/2$ *and a length-scale parameter* $\theta = 0.4$*. In the right panel, we use the HMC sampler (Pakman and Paninski, 2014) to sample from the posterior distribution of* $\{\xi_j\}$ *as in Algorithm 1. The black stars represent the 50 training data generated from (2.1) using the true function* f_1 *and a true noise standard deviation of* $\sigma_{noise} = 0.25$*. First, we observe that both the prediction estimates and the confidence intervals in the left panel do not satisfy boundedness constraints. Second, we observe that including the boundedness constraints into the posterior distribution (right panel) results in more accurate predictions and smaller confidence intervals compared to those produced by the unconstrained GP model (left panel). Let us note that in the left panel, the MAP and mAP estimates coincide in the unconstrained case according to the result in Kimeldorf and Wahba (1970). However, for the constrained case on the right panel, the MAP and mAP no longer coincide. In the present paper, the performance of the MAP estimate is further discussed, with a particular focus on its behavior compared to that of other estimates. Let us conclude this example by noting that, based on this numerical experiment, the MAP estimate appears to perform better visually than the mAP estimate. It is closer to the observed values than the mAP estimate.*

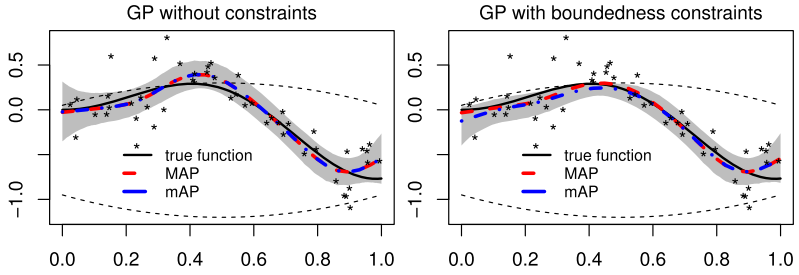

Figure 3: In the left (resp. right) panel, the true function and its estimates (MAP and mAP) are plotted using Model (M_h) without (resp. with) boundedness constraints. The 50 training data are represented by black stars, and the 95% pointwise confidence interval is based on 5,000 McMC iterations and is represented by the gray shaded region. The lower and upper bounds are indicated by the two black dashed thin curves.

Algorithm 1: Sampling scheme of $\{\boldsymbol{\xi}|\boldsymbol{y}, \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\}$, where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Gamma})$.

- Initialization: $\boldsymbol{y}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{N\times N}$, $\mathbf{H} \in \mathbb{R}^{n\times N}$, $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times N}$, $\boldsymbol{\ell}$, $\boldsymbol{u}$.
- Compute the conditional mean and covariance of $\{\boldsymbol{\xi}|\mathbf{H}\boldsymbol{\xi} + \boldsymbol{\epsilon} = \boldsymbol{y}\}$

$$\begin{aligned} \boldsymbol{\mu} &= (\mathbf{H}\boldsymbol{\Gamma})^\top(\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^\top + \sigma^2_{\text{noise}}\mathbf{I}_n)^{-1}\boldsymbol{y}; \\ \boldsymbol{\Sigma} &= \boldsymbol{\Gamma} - (\mathbf{H}\boldsymbol{\Gamma})^\top(\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^\top + \sigma^2_{\text{noise}}\mathbf{I}_n)^{-1}\mathbf{H}\boldsymbol{\Gamma}. \end{aligned}$$

- Compute the posterior mode by solving the quadratic optimization problem subject to linear inequality constraints

$$\boldsymbol{\mu}^* := \arg\min_{\boldsymbol{z}\in\mathbb{R}^N} \{\boldsymbol{z}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{z}|\mathbf{H}\boldsymbol{z} + \boldsymbol{\epsilon} = \boldsymbol{y}, \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{z} \leq \boldsymbol{u}\}.$$

- Sample from the truncated MVN distribution (HMC `tmg` is used in this paper)

$$\{\boldsymbol{\xi}|\mathbf{H}\boldsymbol{\xi} + \boldsymbol{\epsilon} = \boldsymbol{y}, \boldsymbol{\ell} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\} \sim \mathcal{N}_{\mathcal{T}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\ell}, \boldsymbol{u}).$$

Example 2 (Sequential constraints)**.** *Model (M_h) is capable of incorporating different shape constraints sequentially at nonoverlapping intervals as shown in Figure 4 below and Figure 3 of the supplementary material.*

In Figure 4, the lower and upper bounds functions f_ℓ and f_u are

$$f_\ell(x) = (x - 0.5)^2 + 0.1 \quad \textit{and} \quad f_u(x) = \begin{cases} -x + 0.8 & \textit{if } x \in [0, 0.4]; \\ 0.5x + 0.2 & \textit{if } x \in (0.4, 1]. \end{cases}$$

We use the Matérn covariance function with $\nu = 5/2$ and $\theta = 0.4$. The black stars represent the training observations, where $\{x_i\}$ are generated uniformly on $[0, 1]$ and the data $\{y_i\}$ are generated uniformly on $[0, 0.6]$. The input domain is split into two

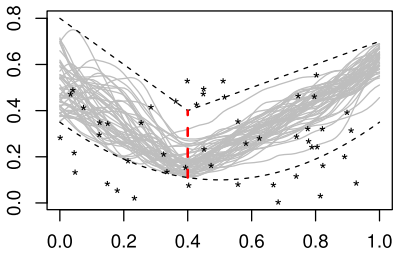

Figure 4: GP approximation from Model (M_h) under boundedness constraints on $[0, 0.4]$ and under both boundedness and monotonicity constraints on $[0.4, 1]$. The vertical dashed line divides the interval $[0, 1]$ into two nonoverlapping subintervals at $x = 0.4$.

nonoverlapping subdomains by a red vertical dashed line at $x = 0.4$. *We suppose that the underlying function admits two different behaviors within these two subdomains. In the first subdomain* $[0, 0.4]$, *the function is bounded. In the second subdomain* $[0.4, 1]$, *it is both nondecreasing and bounded. Again, we applied Proposition 2 with* $Q = 2$, *ensuring that* $x = 0.4$ *is included as an element in the subdivision* $\{t_j\}$. *The gray curves represent the GP sample paths from Model* (M_h) *under sequential constraints. The sample paths respect boundedness constraints on the entire domain and nondecreasing constraints on* $[0.4, 1]$. *It is worth noting that the sequential constraints can be employed to generate GP sample paths that are bounded within a non-convex set (see Example 2 in the supplementary material).*

Example 3 (Impact of outliers)**.** *The aim of this numerical example is to investigate the impact of outliers on the robustness of our model. Robustness, in general terms, is the property of a statistical method to yield sensible results even if its assumptions are violated. In a more specific sense, this refers to the model's insensitivity to outliers (Hampel, 1971, 1974; Huber, 1964). Recently, methods for robustly fitting Gaussian graphical models have been developed in Vogel et al. (2023).*

For our analysis, outlier location crucially affects robustness. In Figure 5, we revisit boundedness in a convex set from Example 1, both with and without outliers. An outlier is added at $x = 0.45$ *(middle panel) and* $x = 1$ *(right panel). The model's prediction*

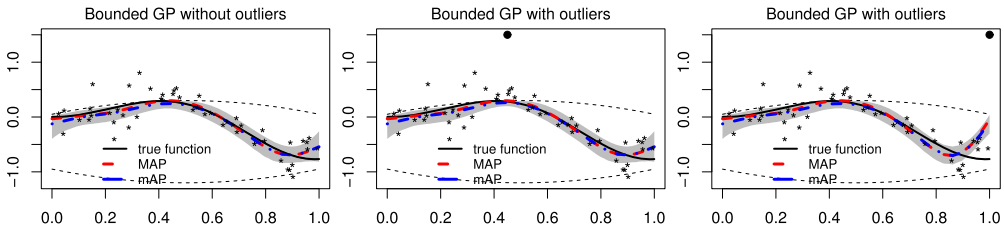

Figure 5: Model (M_h) performance under boundedness constraints is illustrated with data: without outliers (left), with an outlier at $x = 0.45$ (middle), and at $x = 1$ (right).

is notably affected at $x = 1$, but not at $x = 0.45$, where the function already meets the upper bound constraints, unlike at $x = 1$. It is worth noting that the impact of outliers is less significant in locations where the underlying function meets the lower and upper bound constraints. In some situations, the constraints mitigate the effect of outliers.

3.4 Multidimensional input spaces

The finite-dimensional GP approximation defined in (M_h) can be extended to d-dimensional input spaces by tensorization (Maatouk and Bay, 2017). For simplicity of notations, we focus on the case $d = 2$, with $\mathcal{D}^2 = [0, 1]^2$ and $N_1 \times N_2$ knots located on a regular (or non regular) grid. Then, for any $\boldsymbol{x} = (x_1, x_2) \in \mathcal{D}^2$, the finite-dimensional Gaussian approximation is given by

$$Y^{N_1,N_2}(x_1, x_2) := \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \xi_{j_1,j_2} h^1_{j_1}(x_1) h^2_{j_2}(x_2), \tag{3.14}$$

where $\{h^1_{j_1}\}$ and $\{h^2_{j_2}\}$ are the *hat* functions defined in (3.1), and $\xi_{j_1,j_2} = Y(t_{j_1}, t_{j_2})$, with $\{(t_{j_1}, t_{j_2})\}$ the knots. Similarly to the one-dimensional case, $\boldsymbol{\xi} = (\xi_{j_1,j_2}) \in \mathbb{R}^{N_1 \times N_2}$ is a zero-mean Gaussian vector with covariance matrix $\boldsymbol{\Gamma}$ as in (3.2). Thus

- For monotonicity in two dimensions, the constraints to be satisfied are given by $\xi_{j+1,l} \geq \xi_{j,l}$ and $\xi_{j,l+1} \geq \xi_{j,l}$ for any $j \in \{1, \ldots, N_1 - 1\}$ and $l \in \{1, \ldots, N_2 - 1\}$. The constraints for the monotonicity with respect to one of the two input variables can be computed in a similar way (Maatouk and Bay, 2017).

- For convexity in two dimensions, the constraints to be satisfied are given by

$$\frac{\xi_{j+1,l} - \xi_{j,l}}{t_{j+1} - t_j} \leq \frac{\xi_{j+2,l} - \xi_{j+1,l}}{t_{j+2} - t_{j+1}} \text{ and } \frac{\xi_{j,l+1} - \xi_{j,l}}{t_{l+1} - t_l} \leq \frac{\xi_{j,l+2} - \xi_{j,l+1}}{t_{l+2} - t_{l+1}},$$

  for any $j \in \{1, \ldots, N_1 - 2\}$ and $l \in \{1, \ldots, N_2 - 2\}$. The constraints for convexity with respect to one of the two input variables can be computed in a similar way.

- For the upper and lower bound functions, f_u and f_ℓ respectively, that define a convex set, the constraints are expressed as $f_\ell(t_j, t_l) \leq \xi_{j,l} \leq f_u(t_j, t_l)$, for all $j \in \{1, \ldots, N_1\}$ and $l \in \{1, \ldots, N_2\}$.

Remark 2. *The boundedness constraints in the last item above can be extended to the case where the region between the lower and upper bounds is non-convex. As in the one-dimensional case, the only requirement is that the points where the input domain $\mathcal{D}^2$ is divided into convex subsets must be part of the discretization grid (please refer to Example 4 in the supplementary material).*

Example 4 (Numerical illustrations in 2D)**.** *The purpose of this numerical example is to demonstrate the effectiveness of the proposed approach in the two-dimensional scenario. The flexibility of Model (3.14) in incorporating multiple constraints is highlighted.*

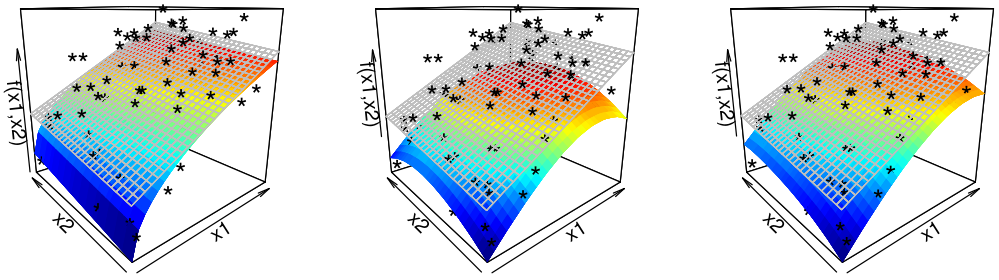

Figure 6: The true nondecreasing and bounded function (left). The MAP estimate from Model (3.14) under boundedness constraints (middle) and under both boundedness and monotonicity constraints (right). The gray surface represents the upper bound constraint, and the black stars represent the training data.

Figure 6 shows an example where boundedness and monotonicity constraints in two dimensions are imposed together. The two-dimensional squared exponential (SE) covariance function is used

$$k(\boldsymbol{x},\boldsymbol{x}') = \exp\left(-\frac{(x_1-x_1')^2}{2\theta_1} - \frac{(x_2-x_2')^2}{2\theta_2}\right), \quad \boldsymbol{x}=(x_1,x_2)\in\mathcal{D}^2, \tag{3.15}$$

where θ_1 and θ_2 are the length-scale hyperparameters. The one hundred training observations (black stars) were generated using the Hypercube Latin from the `R package` *lhs, Eq. (2.1), the function $f(x_1,x_2)=5.6\sqrt{x_1}+x_2+10$, and a true $\sigma_{noise}=1$. This function is monotone nondecreasing with respect to the two input variables and bounded from above by $f_u(x_1,x_2)=4x_1+x_2+12$. In this figure, we illustrate the true function with the upper bound constraints on the left, the MAP estimate with boundedness constraints in the middle, and the MAP with both boundedness and monotonicity constraints on the right. The addition of multiple constraints enhances the accuracy of predictions.*

4 Constrained GPs: C^p approximation, $p \geq 1$

In this section, we generalize Model (M_h) in order to provide smoother sample paths by proposing new basis functions. The capability of this new model to incorporate multiple constraints such as monotonicity with bounded slope constraints is investigated. Furthermore, a comparison of the prediction accuracy between these models is included (Section 5 in the supplementary material).

4.1 C^1 approximation with Model (M_ϕ)

In this section, shape constraints for continuous and differentiable functions $f \in C^1(\mathcal{D},\mathbb{R})$ is considered. For example, the convex set $\mathcal{C}_m$ of nondecreasing functions is given by

$$\mathcal{C}_m = \left\{f \in C^1(\mathcal{D},\mathbb{R}) \text{ s.t. } f'(x)\geq 0,\ x\in\mathcal{D}\right\}.$$

As stated in Maatouk and Bay (2017), any at least differentiable function f can be written as $f(x) = f(0) + \int_0^x f'(t)\mathrm{d}t$, where $f'(t)$ represents the derivative of f at t. Following the strategy of Section 3.1, any differentiable function f can be approximated by $\tilde{f}_N(x) = f(0) + \sum_{j=1}^N f'(t_j)\phi_j(x)$, for any $x \in \mathcal{D}$, where we recall that $\phi_j(x) = \int_0^x h_j(t)\mathrm{d}t$, for any $x \in \mathcal{D}$, with $\{h_j\}$ the hat basis functions defined in (3.1).

Lemma 2 (Uniform convergence C^1)**.** *For any $f \in C^1(\mathcal{D}, \mathbb{R})$, the function $\tilde{f}_N := f(0) + \sum_{j=1}^N f'(t_j)\phi_j(x)$ converges uniformly to f when N tends to infinity.*

In Figure 7 (right panel), the red solid curve is the deterministic function $f(x) = x^3$, which verifies monotonicity (nondecreasing) constraints. The black dashed curve, represented by $\tilde{f}_N(x) = f(0) + \sum_{j=1}^N f'(t_j)\phi_j(x)$, is an approximation of $f(x)$ using a uniform subdivision with $N = 5$ basis functions. The gray curves represent the basis functions $\{\phi_j\}$ defined in (3.1). The black stars represent the value of the true function at knots $\{t_j\}$. In contrast, the left panel shows the approximation of $f'(x) = 3x^2$ (red solid curve) using the derivative of our proposed approach $\tilde{f}'_N$ (black dashed curve) and the hat basis functions $\{h_j\}$ (gray triangles).

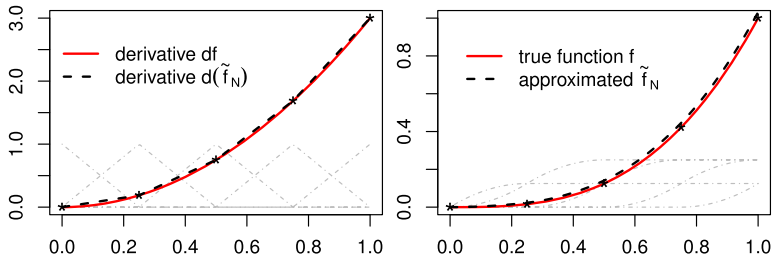

Figure 7: Left: the approximation of $f'(x) = 3x^2$ (red solid curve) by the derivative of the proposed approach $\tilde{f}'_N$ (black dashed curve) together with the hat functions $\{h_j\}$ (gray triangles). Right: the approximation of the nondecreasing function $f(x) = x^3$ (red solid curve) by the proposed approach $\tilde{f}_N$ together with the basis functions $\{\phi_j\}$ (gray curves). A uniform subdivision is used with $N = 5$ basis functions.

Now we consider the second model proposed in Maatouk and Bay (2017)

$$Y^N(x) := Y(0) + \sum_{j=1}^N Y'(t_j)\phi_j(x) = \xi_0 + \sum_{j=1}^N \xi_j\phi_j(x), \quad x \in \mathcal{D}, \qquad (M_\phi)$$

where, we denote by $\xi_j = Y'(t_j)$, for any $j \in \{1, \ldots, N\}$, $\xi_0 = Y(0)$, and $\{\phi_j\}$ are the basis functions defined in (3.1). A comparison between Models (M_h) and (M_ϕ) is presented in Section 5 of the supplementary material. The proposed approach, i.e., Model (M_ϕ) is also applicable for non-uniform subdivision as in Section 3.1. This model has been considered in Zhou et al. (2019) for revisiting the proton-radius problem, and more recently in Zhou et al. (2022) to describe the *mass-shifting* phenomenon of the truncated MVN distribution for a flat region problem. Since $(Y(x))_{x\in\mathcal{D}}$ is a zero-mean

GP, then the vector $[Y(t_1),\ldots,Y(t_N)]^\top$ is also zero-mean, and Gaussian (Rasmussen and Williams, 2006). By Cramer and Leadbetter (1967); Parzen (1962), we know that $[Y'(t_1),\ldots,Y'(t_N)]^\top$ is still a Gaussian vector with covariance matrix

$$\mathbf{G}_{j,l} = \mathrm{Cov}(\xi_j,\xi_l) = \frac{\partial^2}{\partial x \partial x'} k(t_j,t_l), \quad \forall j,l \in \{1,\ldots,N\},$$

where we recall that k is the covariance function of the original GP Y. Therefore, the covariance matrix of the Gaussian vector $\boldsymbol{\xi} = [\xi_0,\xi_1,\ldots,\xi_N]^\top$ is given by

$$\boldsymbol{\Gamma} = \begin{pmatrix} k(0,0) & \frac{\partial}{\partial x'}k(0,t_l) \\ \frac{\partial}{\partial x}k(t_j,0) & \mathbf{G}_{j,l} \end{pmatrix}_{1\le j,l\le N} \in \mathbb{R}^{(N+1)^2}.$$

Thus, $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_{N+1},\boldsymbol{\Gamma})$ as in Section 3.1, where, $\mathbf{0}_{N+1} = [0,\ldots,0]^\top$.

Proposition 3 (Monotonicity and bounded slope)**.** *If the GP approximation Y^N is defined as in (M_ϕ), then, Y^N is nondecreasing (resp. nonincreasing) and $(Y^N)'(x) \in [\ell,u]$, for any $x \in \mathcal{D}$ if and only if $\xi_j \in [\ell,u]$ for any $j \in \{1,\ldots,N\}$, where the lower bound ℓ is nonnegative (resp. the upper bound u is nonpositive). This property can be extended to bounded slope constraints at a subset of $\mathcal{D}$ (see the right panel in Figure 9).*

The linear inequality constraints on the coefficients $\{\xi_j\}$ in Proposition 3 can be written in matrix form as follows

$$\boldsymbol{\ell} \le \boldsymbol{\Lambda}\boldsymbol{\xi} \le \boldsymbol{u},$$

where $\boldsymbol{\ell}$ and $\boldsymbol{u}$ are the m-dimensional vectors representing the lower and upper bounds, and $\boldsymbol{\Lambda}$ is the $m \times (N+1)$ matrix of constraints, with m number of linear constraints. Since the lower bound vector $\boldsymbol{\ell}$ is nonnegative, incorporating monotonicity and bounded slope constraints requires only $m = N$ linear constraints on the basis coefficients $\{\xi_j\}$.

Corollary 1. *If Y^N is defined as in Model (M_ϕ), and f_ℓ and f_u are lower and upper bounds functions such that the region between these two functions is convex, then, $(Y^N)'(x) \in [f_\ell(x),f_u(x)]$ for any $x \in \mathcal{D}$ if and only if $\xi_j \in [f_\ell(t_j),f_u(t_j)]$, for any $j \in \{1,\ldots,N\}$.*

Proof. The proof is similar to the one provided in the first item of Proposition 1 in the supplementary material (Section 1). □

Remark 3. *The result in Corollary 1 can be extended to include bounded slope constraints on subsets of $\mathcal{D}$, as well as non-convex regions between lower and upper bounds functions f_ℓ and f_u, by decomposing the input domain $\mathcal{D}$ into nonoverlapping subdomains with convex regions. As in Proposition 2, the only requirement is to include the intersection extremities of the subdomains in the subdivision $\{t_j\}$.*

Proposition 4 (Multiple constraints (M_ϕ))**.** *If the GP approximation Y^N is defined as in (M_ϕ), then*

- *Monotonicity and boundedness: Y^N is nondecreasing and nonnegative if and only if $\{\xi_j\}$ are nonnegative for any $j \in \{0,\ldots,N\}$.*

- *Convexity:* Y^N *is convex if and only if* $\xi_j \leq \xi_{j+1}$, *for any* $j \in \{1, \ldots, N-1\}$.
- *Monotonicity and convexity:* Y^N *is monotone (nondecreasing) and convex in the entire domain if and only if* $0 \leq \xi_1 \leq \ldots \leq \xi_N$.

  This leads to only $m = N$ *linear constraints on the basis coefficients* $\{\xi_j\}$ *according to the notation in* (3.6).
- *Monotonicity, convexity and boundedness:* Y^N *is nondecreasing, convex, and non-negative if and only if* $0 \leq \xi_1 \leq \ldots \leq \xi_N$ *and* $\xi_0 \geq 0$.

Corollary 2. *The bounded slope constraints from Proposition 3 can be combined with those presented in Proposition 4.*

4.2 C^p approximation, $p \geq 2$ with Model (M_ψ)

The sample paths generated from Model (M_ϕ) are differentiable. This means that the derivatives of order zero and one are continuous. This is because the basis functions $\{\phi_j\}$ are the primitive of the *hat* functions $\{h_j\}$ which are only continuous. The differentiability of the sample paths generated from (M_ϕ) can be generalized to any class C^p, $p \geq 1$. For example, to obtain sample paths that are twice differentiable, it is sufficient to define a *hat* basis function that is differentiable, as follows:

$$\kappa_j(x) = \kappa\left(\frac{x - t_j}{\Delta_N}\right) \quad \text{and} \quad \psi_j(x) = \int_0^x \kappa_j(t)\mathrm{d}t, \text{ where} \tag{4.1}$$

$$\kappa(x) = \begin{cases} -2x^3 - 3x^2 + 1 & \text{if } x \in [-1, 0]; \\ 2x^3 - 3x^2 + 1 & \text{if } x \in (0, 1]; \\ 0 & \text{otherwise;} \end{cases} \tag{4.2}$$

$t_j = (j-1) \times \Delta_N$, with $\Delta_N = 1/(N-1)$. The function κ can be seen as the cubic Hermite spline defined on $[-1, 1]$. It is clear that κ is a differentiable function on $\mathbb{R}$ and that $\{\kappa_j\}$ is also differentiable. This implies that $\{\psi_j\}$ are twice differentiable. Additionally, $\kappa_j(t_l) = \delta_{j,l}$, where $\delta_{j,l}$ is the Kronecker's delta function equal to one if $j = l$ and zero otherwise. Furthermore, the 'new' *hat* functions $\{\kappa_j\}$ admit the following nice property $\sum_{j=1}^N \kappa_j(x) = 1$, for any $x \in \mathcal{D}$. This property plays an important role in the bounded slope constraints that are added to the proposed approach, as well as in the convexity constraints (see Proposition 6 below). Following the strategy of Section 4, any differentiable function f can be approximated by

$$\tilde{f}_N(x) = f(0) + \sum_{j=1}^{N} f'(t_j)\psi_j(x), \tag{4.3}$$

for any $x \in \mathcal{D}$.

In Figure 8, a uniform subdivision of the domain $\mathcal{D} = [0, 1]$ is used with $N = 5$ basis functions and knots. The non-uniform subdivision case is straightforward. In the left panel, we show the 'new' *hat* functions $\{\kappa_j\}$ as well as the derivative function

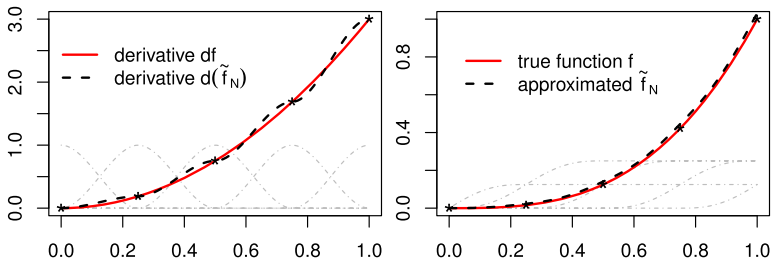

Figure 8: A uniform subdivision is used with $N = 5$ basis functions and knots. Left: $f'(x) = 3x^2$ (red solid curve) and $\tilde{f}'_N$ (black dashed curve) together with the 'new' *hat* functions $\{\kappa_j\}$ (gray curves). Right: the monotone nondecreasing function $f(x) = x^3$ (red solid curve) and the proposed approach $\tilde{f}_N$ in (4.3) represented by the black dashed curve, together with the 'new' basis functions $\{\psi_j\}$ (gray curves).

$f'(x) = 3x^2$ approximated by $\tilde{f}'_N(x) = \sum_{j=1}^N f'(t_j)\kappa_j(x)$. The right panel shows the deterministic function $f(x) = x^3$ (red solid curve), which verifies monotonicity (non-decreasing) constraints and the proposed approach $\tilde{f}_N(x) = f(0) + \sum_{j=1}^N f'(t_j)\psi_j(x)$ (black dashed curve). The gray curves are the basis functions $\{\psi_j\}$ defined in (4.1). The use of the basis functions $\{\psi_j\}$ as in (M_ψ) leads to differentiable GP sample paths of class C^2. The slope of the sample paths can also be controlled, as show in Proposition 6.

Proposition 5 (Bounded slope C^2)**.** *Let f be a continuous and differentiable function on $\mathcal{D}$, i.e., $f \in C^1(\mathcal{D}, \mathbb{R})$, and $\tilde{f}_N(x) := f(0) + \sum_{j=1}^N f'(t_j)\psi_j(x)$, for any $x \in \mathcal{D}$. Then*

$$\tilde{f}'_N(x) \in [\ell, u],\ \forall x \in \mathcal{D} \quad \Leftrightarrow \quad f'(t_j) \in [\ell, u],\ \forall j \in \{1, \dots, N\},$$

where ℓ and u are the lower and upper bounds respectively.

In that case, the proposed approach is defined as follows:

$$Y^N(x) := Y(0) + \sum_{j=1}^N Y'(t_j)\psi_j(x) = \xi_0 + \sum_{j=1}^N \xi_j\psi_j(x), \quad \forall x \in \mathcal{D}, \qquad (M_\psi)$$

where, the basis $\{\psi_j\}$ are defined in (4.1), $\xi_0 = Y(0)$ and $\xi_j = Y'(t_j)$, $j \in \{1, \dots, N\}$.

Proposition 6 (Multiple constraints (M_ψ))**.** *If Y^N is defined as in (M_ψ), then*

- *the sample paths generated from Y^N are twice differentiable.*
- *Monotonicity: $Y^N \in \mathcal{C}_m$ if and only if $\{\xi_j\}$ are nonnegative, for any $j \in \{1, \dots, N\}$.*
- *Monotonicity and bounded slope: $Y^N \in \mathcal{C}_m$ and $(Y^N)'(x) \in [\ell, u]$, for any $x \in \mathcal{D}$ if and only if $\{\xi_j\}$ are in $[\ell, u]$, for any $j \in \{1, \dots, N\}$, where $\ell, u \in \mathbb{R}_+$ are the lower and upper bounds respectively.*

- *Monotonicity and boundedness:* Y^N *is nondecreasing and nonnegative if and only if* $\{\xi_j\}$ *are nonnegative, for any* $j \in \{0, \ldots, N\}$.
- *Convexity:* $Y^N \in \mathcal{C}_c$ *if and only if* $\xi_j \leq \xi_{j+1}$, *for any* $j \in \{1, \ldots, N-1\}$.
- *Monotonicity and convexity:* $Y^N \in \mathcal{C}_m \cap \mathcal{C}_c$ *if and only if* $0 \leq \xi_1 \leq \ldots \leq \xi_N$.
- *Monotonicity, convexity and boundedness:* Y^N *is nondecreasing, convex, and nonnegative if and only if* $0 \leq \xi_1 \leq \ldots \leq \xi_N$ *and* $\xi_0 \geq 0$.

Remark 4. *In this section, the smoothness of the sample paths of Model* (M_ϕ) *has been investigated, with a focus on its differentiability in class* C^p, *for* $p \geq 1$. *The function* κ, *as given in* (4.2), *has a differentiability of class* C^1. *This implies that the sample paths from* (M_ψ) *are twice differentiable. For example, to obtain sample paths that are differentiable up to order three (i.e., class* C^3*), it is sufficient to define* κ *as follows*

$$\kappa(x) = \begin{cases} -3x^4 - 8x^3 - 6x^2 + 1 & \text{if } x \in [-1, 0]; \\ -3x^4 + 8x^3 - 6x^2 + 1 & \text{if } x \in (0, 1]; \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

It is straightforward to verify that κ *is a differentiable function of class* C^2. *Adopting the same approach, we define* κ_j *and* ψ_j *as follows:*

$$\kappa_j(x) = \kappa(x - t_j/\Delta_N) \quad \text{and} \quad \psi_j(x) = \int_0^x \kappa_j(t)\mathrm{d}t.$$

It is worth noting that the basis $\{\kappa_j\}$ *verify the following two fundamental properties:*

$$\kappa_j(t_l) = \delta_{j,l} \quad \text{and} \quad \sum_{j=1}^{N} \kappa_j(x) = 1,$$

for all $j, l = 1, \ldots, N$ *and for any* $x \in \mathcal{D}$, *where we recall that* $\delta_{j,l}$ *is the Kronecker's delta function equal to one if* $j = l$ *and zero otherwise.*

4.3 C^p approximation, $p \geq 2$ with Model (M_φ)

In this section, we first consider the convexity constraint for continuous and twice differentiable functions $f \in C^2$. Thus, the convex set $\mathcal{C}_c$ is given by

$$\mathcal{C}_c = \left\{ f \in C^2(\mathcal{D}, \mathbb{R}) \text{ s.t. } f''(x) \geq 0, \ x \in \mathcal{D} \right\},$$

where f'' represents the second-order derivative of f. As stated in Maatouk and Bay (2017), any at least twice differentiable function f can be written as $f(x) = f(0) + xf'(0) + \int_0^x \int_0^t f''(u)\mathrm{d}u\mathrm{d}t$. Following the strategy of Section 3.1, any twice differentiable function f can be approximated by

$$\tilde{f}_N(x) = f(0) + xf'(0) + \sum_{j=1}^{N} f''(t_j)\varphi_j(x) \tag{4.5}$$

for any $x \in \mathcal{D}$, where $\{\varphi_j\}$ are given in (3.1).

Lemma 3 (Uniform convergence C^2)**.** *For any $f \in C^2(\mathcal{D}, \mathbb{R})$, the function $\tilde{f}_N$ defined in (4.5) converges uniformly to f when N tends to infinity.*

In that case, we consider the following model

$$Y^N(x) := Y(0) + Y'(0)x + \sum_{j=1}^{N} Y''(t_j)\varphi_j(x) = \xi_0^* + \xi_0 x + \sum_{j=1}^{N} \xi_j \varphi_j(x), \quad x \in \mathcal{D}, \quad (M_\varphi)$$

where we denote by $\xi_0^* = Y(0)$, $\xi_0 = Y'(0)$, and $\xi_j = Y''(t_j)$, for any $j \in \{1, \ldots, N\}$. In this case, Y^N is convex on $\mathcal{D}$ *if and only if* $\{\xi_j\}$ are nonnegative for any $j \in \{1, \ldots, N\}$.

Proposition 7 (Multiple constraints (M_φ))**.** *If Y^N is defined as in (M_φ), then*

- *Monotonicity and convexity: $Y^N \in \mathcal{C}_m \cap \mathcal{C}_c$ if and only if $\xi_j \geq 0$ for any $j \in \{0, \ldots, N\}$.*
- *Monotonicity, convexity and boundedness: $Y^N \in \mathcal{C}_m \cap \mathcal{C}_c$ and Y^N is nonnegative if and only if $\xi_0^* \geq 0$ and $\xi_j \geq 0$, for any $j \in \{0, \ldots, N\}$.*

Remark 5. *The sample paths generated from Model (M_φ) are twice differentiable. As in Section 4.2, the smoothness of the sample paths can be generalized to class C^p, for any $p \geq 2$ by defining smoother hat functions like ones given in (4.1) and (4.4).*

4.4 Performance of the MAP estimate

The aim of this section is to show the performance of the MAP estimate in terms of prediction accuracy in different situations. A variety of functions is considered:

$$f_{m_1}(x) = (5x-3)^3 \mathbf{1}_{[0.6,1]}(x), \qquad f_{m_2}(x) = \frac{3}{1+\exp\{-10x+2.1\}},$$
$$f_{m_3}(x) = \sqrt{2}\sum_{j=1}^{100} j^{-1.7}\sin(j) \times \cos(\pi(j-0.5)(1-x)), \qquad f_{m_4}(x) = 5x^2, \quad (4.6)$$

for $x \in [0,1]$. The function f_{m_1} is monotone (nondecreasing) and flat on $[0, 0.6]$. However, f_{m_2} and f_{m_3} are approximately flat on $[0.7, 1]$. The last function f_{m_4} is an increasing function on the whole domain $[0,1]$. Let us mention that only the function f_{m_3} is decreasing in certain regions which allows us to evaluate the performance of the proposed approach under slight model misspecification.

The simulation studies are based on a dataset of size $n = 500$ generated from (2.1) using the true functions (4.6) and a true $\sigma_{\text{noise}} = 0.5$. The dataset is randomly split into training set of size 300 and testing set of size 200. Table 1 summarizes the average of the MSPE $\times 10^2$ (standard deviation $\times 10^2$) over one thousand replicates for the four true functions (4.6) using different approaches MAP estimate, dependent global local prior (DGL), independent global-local shrinkage prior (IGL), and the global shrinkage prior (TMVN). To evaluate the performance between the flat and increasing regions

functions	methods	MSPE (total)	MSPE (flat)	MSPE (increasing)
f_{m_1}	MAP	13.55 (4.97)	**4.39** (1.87)	19.27 (7.59)
	DGL	**11.36** (2.62)	8.13 (1.95)	**14.71** (4.79)
	IGL	13.44 (2.62)	9.86 (1.70)	17.32 (5.23)
	TMVN	65.63 (7.21)	14.53(2.59)	102.6 (11.16)
f_{m_2}	MAP	**7.46** (1.65)	**3.94** (2.31)	8.63 (1.99)
	DGL	8.29 (1.78)	7.13 (2.64)	8.56 (2.32)
	IGL	9.55 (1.92)	8.40 (2.61)	9.84 (2.54)
	TMVN	8.32 (2.11)	8.61 (2.91)	**7.94** (2.75)
f_{m_3}	MAP	7.84 (1.47)	**5.23** (2.23)	8.78 (1.79)
	DGL	7.76 (1.74)	9.16 (2.9)	**6.87** (1.87)
	IGL	**7.72** (1.74)	8.57 (2.45)	7.18 (1.74)
	TMVN	11.36 (1.33)	15.27 (2.85)	8.97 (1.76)
f_{m_4}	MAP	9.44 (1.89)	–	9.44 (1.89)
	DGL	8.67 (2.15)	–	8.67 (2.15)
	IGL	9.34 (2.16)	–	9.34 (2.16)
	TMVN	**5.68** (1.61)	–	**5.68** (1.61)

Table 1: The average MSPE $\times 10^2$ (standard deviation $\times 10^2$) over one thousand replicates for different functions and methods. The MAP estimate is obtained through Model (M_h), under only monotonicity constraints.

separately, we additionally report the average partial MSPEs for each region: MSPE (flat) for the flat portion and MSPE (increasing) for the increasing portion, in addition to the overall average MSPE. To avoid overfitting, we set $N = \lfloor n_{tr}/8 \rfloor$ as in Zhou et al. (2022), where n_{tr} is the number of training samples fixed at 300. In that case, the MAP estimate provides the same MSPE results when using different models: (M_h), (M_ϕ), and (M_φ) (see Section 5 in the supplementary material). The Matérn family of covariance functions is used with smoothness parameter $\nu \sim \mathcal{U}(0.5, 1)$ and length-scale parameter $\theta \sim \mathcal{U}(0.1, 1)$ generated at each replicate as in Zhou et al. (2022). For flat regions, the MAP estimate outperformed the shrinkage approaches of Zhou et al. (2022) (DGL, IGL, and TMVN). This confirms the robustness of the MAP estimate for capturing flat regions. According to the MSPE criterion, the MAP estimate is twice (resp. three times) more efficient than IGL and DGL (resp. TMVN) when using f_{m_2} over the flat region. This was also seen when calculating the total MSPE for f_{m_2}, where the proposed approach had a slightly lower standard deviation than the shrinkage approaches (IGL, DGL, and TMVN). This again confirms the stability of the MAP estimate provided by the proposed approach. Finally, it is worth noting that the simulation studies are conducted without any additional constraints.

Figure 9 shows the finite-dimensional GP approximation from Model (M_ϕ) with monotonicity constraints only (left panel) and both monotonicity and bounded slope constraints (right panel). The function $f_{m_2}(x) = \frac{3}{1+\exp\{-10x+2.1\}}$ is considered. This function is interesting because it is monotonically increasing and almost flat over the interval $[0.7, 1]$. We used the Matérn covariance function with $\nu = 5/2$ and the efficient HMC technique (Pakman and Paninski, 2014) to sample from the posterior distribution

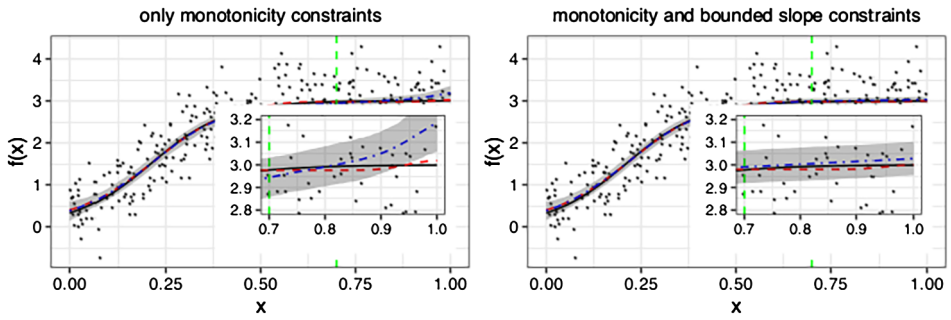

Figure 9: GP approximation from Model (M_ϕ) satisfying monotonicity (left panel), monotonicity and bounded slope constraints (right panel). The panel description is the same as in Figure 3, with zoomed-in inset plots where $x \in [0.7, 1]$. The green dashed line in the right panel represents the starting location of bounded slope constraints.

of the basis coefficients $\{\xi_j\}$. The black solid curve represents the function f_{m_2}, while the red dashed (resp. blue dashed-dotted) curve represents the MAP estimate (resp. mAP estimate). The gray shaded area represents the 95% pointwise confidence interval. The black stars are the 300 noisy observations generated from (2.1) using f_{m_2} and a true noise standard deviation of $\sigma_{\text{noise}} = 0.5$. The green vertical dashed line in the right panel corresponds to the starting point where the bounded slope constraints are imposed. In the right panel, we impose an upper bound slope constraints on the proposed approach in the flat region $[0.7, 1]$. In the left panel, we observe that the pointwise 95% confidence intervals fail to capture the true nondecreasing function f_{m_2} for a substantial part of the input domain. This is due to the *mass-shifting* phenomenon highlighted in Zhou et al. (2022). Including bounded slope constraints (right panel) provides smaller and more realistic credible intervals as compared to those without such constraints (left panel). Let us mention that the MAP estimate is robust in both scenarios, with or without bounded slope constraints. This is because the MAP estimate converges to the constrained optimal smoothing function (as proved in Grammont et al. (2022)). Additionally, the mAP estimate tends towards the MAP estimate when bounded slope constraints are added, as seen in the right panel. Finally, the average 95% posterior coverage over 25 replicates is equal to 78% when using the proposed approach with only monotonicity (nondecreasing) constraints. However, it increases to 89% when adding the bounded slope constraints. Thus, adding multiple constraints improves prediction accuracy, bringing the posterior coverage closer to 95%.

5 Real-world data applications

5.1 Light detection and ranging (LiDAR)

In this section, the proposed approach developed in this paper was applied on the light detection and ranging (LiDAR) real-world data that consist of 221 observations from a LiDAR experiment and it contain information on range and logratio. The predictor range

represents the distance travelled before the light is reflected back to its source, however, the response variable logratio represents the logarithm of the ratio of received light from two laser sources. This real-world data is available from the `R package` *HRW*. The data suggest that the underlying function is nonpositive and monotone nonincreasing with a flat region when the range is less than 550.

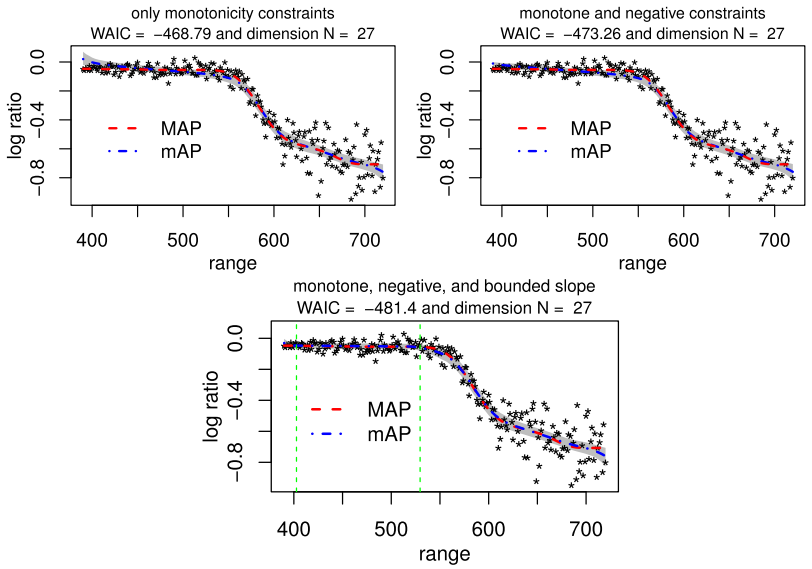

Figure 10: Estimation accuracy of the proposed approach applied on LiDAR data.

In Figure 10, Model (M_ϕ) has been applied to the $n = 221$ LiDAR data (shown as black stars). We fix $N = \lfloor n/8 \rfloor$ to avoid overfitting (Section 6.1 of the supplementary material justifies this choice), and we use the Matérn covariance functions with $\nu = 5/2$. Top left: Model (M_ϕ) satisfying monotonicity constraints only. We observe that, unlike the MAP estimate, the credible interval and mAP estimate fail to follow the behavior of the data in the flat region (range less than 550). To rigorously compare the MAP and mAP estimates in terms of prediction accuracy, we propose to place a uniform prior distribution on the length-scale parameter $\theta \sim \mathcal{U}(50, 300)$ as well as on the noise standard deviation $\sigma_{\text{noise}} \sim \mathcal{U}(0.1, 0.5)$. By randomly splitting the total dataset of size 221 into 80% training and 20% testing datasets, we obtain an average MSPE over one thousand replicates of 8.23×10^{-2} when using the MAP estimate and 9.76×10^{-2} when using the mAP estimate. Top right: we added nonnegativity constraints (as per Proposition 4) and found that, once again, credible intervals as well as the mAP estimate fail to capture the flat region, specially when logratio starts to decreases. Bottom: we added bounded slope constraints in the flat region between the green vertical dashed lines (see Proposition 3) to monotonicity and negativity constraints. In that case, we observe that both the mAP estimate and the 95% pointwise credible interval follow the observations and capture the flat region. The proposed model with triple inequality constraints seems to align with the data better, specifically over the *flat* region and when

logratio starts to decreases. We also observe that the mAP estimate tends towards the MAP estimate, which behaves well in all three situations. As expected, adding multiple constraints improves the prediction accuracy of the proposed approach (MAP and mAP estimates) as well as the behavior of the posterior distribution, which provides more realistic pointwise credible intervals. To provide more precise information, we present the WAIC (Watanabe-Akaike Information Criterion) values for overall model fitting in the main of each panel. As expected, the proposed approach under triple constraints, achieves the lowest WAIC value, indicating a better fit to the data compared to the other two scenarios.

	type constraints	length-scale parameter	noise sd	WAIC
Top left	only monotone	$\theta = 100$	$\sigma_{\text{noise}} = 0.1$	-468.79
Top right	monotone/negative	$\theta = 100$	$\sigma_{\text{noise}} = 0.1$	-473.26
Bottom	monotone/negative bounded slope	$\theta = 100$	$\sigma_{\text{noise}} = 0.1$	-481.4

Table 2: Performance of the proposed approach Model (M_ϕ) under multiple constraints.

Table 2 summarizes the values of hyperparameters θ and σ_{noise}, and the WAIC values using Model (M_ϕ) across various scenarios for LiDAR data. Model (M_ϕ), when applied under constraints of monotonicity, negativity, and bounded slope, provides the smallest WAIC value. This indicates a better fit to the data compared to the other scenarios.

5.2 Nuclear safety application

For brevity, this real application has been deferred to the supplementary material (Section 6.2).

Conclusion

The finite-dimensional Gaussian process approximation originally proposed in Maatouk and Bay (2017) is considered, which verifies interpolation conditions and inequality constraints in the entire domain. The flexibility of this approach to incorporate both noisy observations and multiple convex and non-convex constraints is investigated. This leads to significant improvement in prediction accuracy and more realistic credible intervals. We propose an adjustment to the cross-validation technique that uses *Maximum a Posteriori* (MAP) to estimate both covariance and noise variance hyperparameters. Additionally, we propose new basis functions to enhance the smoothness of the sample paths and ensure differentiability of class C^p, for any $p \geq 1$. The behavior of this approach in challenging situations, such as monotonicity with a flat region or boundedness where the underlying function is flat and close to lower and/or upper bounds, is investigated. In that case, we show that, unlike the MAP estimate, the truncated multivariate normal distribution is not suitable for capturing the flat region. To address this issue, we propose adding multiple constraints, such as monotonicity with bounded slope constraints. The superiority of the MAP estimate over the mean a posterior (mAP)

estimate is demonstrated in a wide range of settings based on its theoretical convergence. Real-world data studies show that the MAP estimate effectively captures flat regions and that incorporating multiple constraints accurately reflects the behavior of the posterior distribution.

Acknowledgments

The authors express gratitude to the Editor, the Associate Editor, and the two reviewers for their valuable and constructive comments, significantly enhancing the presentation and accuracy of the manuscript. The authors would also express their gratitude to Yann Richet from the Institut of Radioprotection and Nuclear Safety (IRSN, Paris) for providing the nuclear safety data. This research was conducted with the support of the consortium in Applied Mathematics CIROQUO, gathering partners in technological and academia in the development of advanced methods for Computer Experiments. https://doi.org/10.5281/zenodo.6581217

Supplementary Material

Supplement to "Bayesian Analysis of Constrained Gaussian Processes" (DOI: 10.1214/24-BA1429SUPP; .pdf). For clarity of the paper, both the theoretical proofs and some numerical simulations and results have been moved to the supplementary material.

References

Bay, X., Grammont, L., and Maatouk, H. (2016). "Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation." *Electronic Journal of Statistics*, 10(1): 1580–1595. MR3507374. doi: https://doi.org/10.1214/16-EJS1149. 2, 10

Botev, Z. I. (2017). "The normal law under linear restrictions: simulation and estimation via minimax tilting." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1): 125–148. MR3597967. doi: https://doi.org/10.1111/rssb.12162. 9

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press. MR2061575. doi: https://doi.org/10.1017/CBO9780511804441. 9

Cai, B. and Dunson, D. B. (2007). "Bayesian multivariate isotonic regression splines: applications to carcinogenicity studies." *Journal of the American Statistical Association*, 102(480): 1158–1171. MR2412540. doi: https://doi.org/10.1198/016214506000000942. 2, 6

Chataigner, M., Cousin, A., Crépey, S., Dixon, M., and Gueye, D. (2021). "Beyond surrogate modeling: Learning the local volatility via shape constraints." *SIAM Journal on Financial Mathematics*, 12(3): SC58–SC69. MR4301403. doi: https://doi.org/10.1137/20M1381538. 2

Chipman, H. A., George, E. I., McCulloch, R. E., and Shively, T. S. (2022). "mBART: Multidimensional Monotone BART." *Bayesian Analysis*, 17(2): 515 – 544. MR4483229. doi: https://doi.org/10.1214/21-ba1259. 2

Cong, Y., Chen, B., and Zhou, M. (2017). "Fast simulation of hyperplane-truncated multivariate normal distributions." *Bayesian Analysis*, 12(4): 1017 – 1037. MR3724977. doi: https://doi.org/10.1214/17-BA1052. 9

Cousin, A., Deleplace, A., and Misko, A. (2022). "Gaussian Process Regression for Swaption Cube Construction under No-Arbitrage Constraints." *Risks*, 10(12): 232. 2

Cousin, A., Maatouk, H., and Rullière, D. (2016). "Kriging of financial term-structures." *European Journal of Operational Research*, 255(2): 631–648. MR3518828. doi: https://doi.org/10.1016/j.ejor.2016.05.057. 2

Cramer, H. and Leadbetter, R. (1967). *Stationary and related stochastic processes: sample function properties and their applications*. Wiley series in probability and mathematical statistics. Tracts on probability and statistics. Wiley. MR0217860. 2, 17

Crépey, S. and Dixon, M. F. (2020). "Gaussian process regression for derivative portfolio modeling and application to credit valuation adjustment computations." *Journal of Computational Finance*, 24(1). 2

Curtis, S. M. and Ghosh, S. K. (2011). "A variable selection approach to monotonic regression with Bernstein polynomials." *Journal of Applied Statistics*, 38(5): 961–976. MR2782409. doi: https://doi.org/10.1080/02664761003692423. 2, 6

Golchi, S., Bingham, D. R., Chipman, H., and Campbell, D. A. (2015). "Monotone Emulation of Computer Experiments." *SIAM/ASA Journal on Uncertainty Quantification*, 3(1): 370–392. MR3349045. doi: https://doi.org/10.1137/140976741. 2

Goldfarb, D. and Idnani, A. (1983). "A numerically stable dual method for solving strictly convex quadratic programs." *Mathematical Programming*, 27(1): 1–33. MR0712108. doi: https://doi.org/10.1007/BF02591962. 9

Grammont, L., Maatouk, H., and Bay, X. (2022). "Equivalent between constrained optimal smoothing and Bayesian estimation." Working paper or preprint. 2, 10, 23

Hampel, F. R. (1971). "A General Qualitative Definition of Robustness." *The Annals of Mathematical Statistics*, 42(6): 1887 – 1896. MR0301858. doi: https://doi.org/10.1214/aoms/1177693054. 13

Hampel, F. R. (1974). "The Influence Curve and Its Role in Robust Estimation." *Journal of the American Statistical Association*, 69(346): 383–393. MR0362657. 13

Huber, P. J. (1964). "Robust estimation of a location parameter." *The Annals of Mathematical Statistics*, 35(1): 73 – 101. MR0161415. doi: https://doi.org/10.1214/aoms/1177703732. 13

Kimeldorf, G. and Wahba, G. (1970). "A correspondence between Bayesian estimation on stochastic processes and smoothing by splines." *The Annals of Mathematical Statistics*, 495–502. MR0254999. doi: https://doi.org/10.1214/aoms/1177697089. 2, 10, 11

Lenk, P. J. and Choi, T. (2017). "Bayesian analysis of shape-restricted functions using Gaussian process priors." *Statistica Sinica*, 43–69. MR3642448. 2

Lin, L. and Dunson, D. B. (2014). "Bayesian monotone regression using Gaussian process projection." *Biometrika*, 101(2): 303–317. MR3215349. doi: https://doi.org/10.1093/biomet/ast063. 2

López-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2018). "Finite-dimensional Gaussian approximation with linear inequality constraints." *SIAM/ASA Journal on Uncertainty Quantification*, 6(3): 1224–1255. MR3857898. doi: https://doi.org/10.1137/17M1153157. 2

Maatouk, H. (2022). "Finite-dimensional approximation of Gaussian processes with linear inequality constraints and noisy observations." *Communications in Statistics-Theory and Methods*, 1–20. MR4649225. doi: https://doi.org/10.1080/03610926.2022.2055768. 8

Maatouk, H. and Bay, X. (2017). "Gaussian process emulators for computer experiments with inequality constraints." *Mathematical Geosciences*, 49(5): 557–582. MR3660323. doi: https://doi.org/10.1007/s11004-017-9673-2. 1, 2, 3, 4, 6, 7, 9, 14, 16, 20, 25

Maatouk, H., Bay, X., and Rullière, D. (2022). "A note on simulating hyperplane-truncated multivariate normal distributions." *Statistics & Probability Letters*, 191: 109650. MR4473831. doi: https://doi.org/10.1016/j.spl.2022.109650. 9

Maatouk, H., Rullière, D., and Bay, X. (2023a). "Large scale Gaussian processes with Matheron's update rule and Karhunen-Loève expansion." In *To appear in: A. Hinrichs, P. Kritzer, F. Pillichshammer (eds.). Monte Carlo and Quasi-Monte Carlo Methods 2022*. Springer Verlag. 9

Maatouk, H., Rullière, D., and Bay, X. (2023b). "Sampling large hyperplane-truncated multivariate normal distributions." *Computational Statistics*, 1–28. 9

Maatouk, H., Rullière, D., and Bay, X., (2024). "Supplementary Material for "Bayesian Analysis of Constrained Gaussian Processes"." doi: https://doi.org/10.1214/24-BA1429SUPP. 10

Maroñas, J., Hamelijnck, O., Knoblauch, J., and Damoulas, T. (2021). "Transforming Gaussian processes with normalizing flows." In *International Conference on Artificial Intelligence and Statistics*, 1081–1089. PMLR. 2

Maroñas, J. and Hernández-Lobato, D. (2023). "Efficient transformed Gaussian processes for non-stationary dependent multi-class classification." In *International Conference on Machine Learning*, 24045–24081. PMLR. 2

Meyer, M. C., Hackstadt, A. J., and Hoeting, J. A. (2011). "Bayesian estimation and inference for generalised partial linear models using shape-restricted splines." *Journal of Nonparametric Statistics*, 23(4): 867–884. MR2854243. doi: https://doi.org/10.1080/10485252.2011.597852. 2, 6

Murphy, K. P. (2018). "Machine learning: A probabilistic perspective (adaptive computation and machine learning series)." 9

Neelon, B. and Dunson, D. B. (2004). "Bayesian isotonic regression and trend analysis." *Biometrics*, 60(2): 398–406. MR2066274. doi: https://doi.org/10.1111/j.0006-341X.2004.00184.x. 11

Pakman, A. and Paninski, L. (2014). "Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians." *Journal of Computational and Graphical Statistics*, 23(2): 518–542. MR3215823. doi: https://doi.org/10.1080/10618600.2013.788448. 9, 11, 22

Parzen, E. (1962). *Stochastic processes*. Holden-Day series in probability and statistics. San Francisco, London, Amsterdam: Holden-Day. MR0139192. 17

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge. MR2514435. 1, 2, 3, 4, 9, 17

Ray, P., Pati, D., and Bhattacharya, A. (2020). "Efficient Bayesian shape-restricted function estimation with constrained Gaussian process priors." *Statistics and Computing*, 30(4): 839–853. MR4108680. doi: https://doi.org/10.1007/s11222-020-09922-0. 2

Riihimäki, J. and Vehtari, A. (2010). "Gaussian processes with monotonicity information." In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 645–652. JMLR Workshop and Conference Proceedings. 2

Shively, T. S., Walker, S. G., and Damien, P. (2011). "Nonparametric function estimation subject to monotonicity, convexity and other shape constraints." *Journal of Econometrics*, 161(2): 166–181. MR2774935. doi: https://doi.org/10.1016/j.jeconom.2010.12.001. 2, 6

Swiler, L. P., Gulian, M., Frankel, A. L., Safta, C., and Jakeman, J. D. (2020). "A survey of constrained Gaussian process regression: Approaches and implementation challenges." *Journal of Machine Learning for Modeling and Computing*, 1(2). 2

Taylor, J. and Benjamini, Y. (2016). "RestrictedMVN: multivariate normal restricted by affine constraints." *R package version*, 1. 9

Tran, T.-T., Fradi, A., and Samir, C. (2023). "Learning, inference, and prediction on probability density functions with constrained Gaussian processes." *Information Sciences*, 642: 119068. 2

Ustyuzhaninov, I., Kazlauskaite, I., Ek, C. H., and Campbell, N. (2020). "Monotonic Gaussian process flows." In *International Conference on Artificial Intelligence and Statistics*, 3057–3067. PMLR. 2

Vogel, D., Watt, S. J., and Wiedemann, A. (2023). *Robustly Fitting Gaussian Graphical Models—the R Package robFitConGraph*, 277–296. Cham: Springer International Publishing. MR4633227. doi: https://doi.org/10.1007/978-3-031-22687-8_13. 13

Wang, X. and Berger, J. O. (2016). "Estimating shape constrained functions using Gaussian processes." *SIAM/ASA Journal on Uncertainty Quantification*, 4(1): 1–25. MR3452261. doi: https://doi.org/10.1137/140955033. 2

Williams, N. J., Osborne, C., Seymour, I. D., Bazant, M. Z., and Skinner, S. J. (2023). "Application of finite Gaussian process distribution of relaxation times on SOFC electrodes." *Electrochemistry Communications*, 107458. 2

Zhou, S., Giulani, P., Piekarewicz, J., Bhattacharya, A., and Pati, D. (2019). "Reexamining the proton-radius problem using constrained Gaussian processes." *Physical Review C*, 99: 055202. 2, 16

Zhou, S., Ray, P., Pati, D., and Bhattacharya, A. (2022). "A mass-shifting phenomenon of truncated multivariate normal priors." *Journal of the American Statistical Association*, 0(ja): 1–37. MR4713915. doi: https://doi.org/10.1080/01621459.2022.2129059. 2, 16, 22, 23